Computational statistics: a silent revolution

Balázs Kégl

AppStat: Machine Learning and Applied Statistics group

Linear Accelerator Laboratory / CNRS & University Paris Sud

Clouds pour le Calcul Scientifique

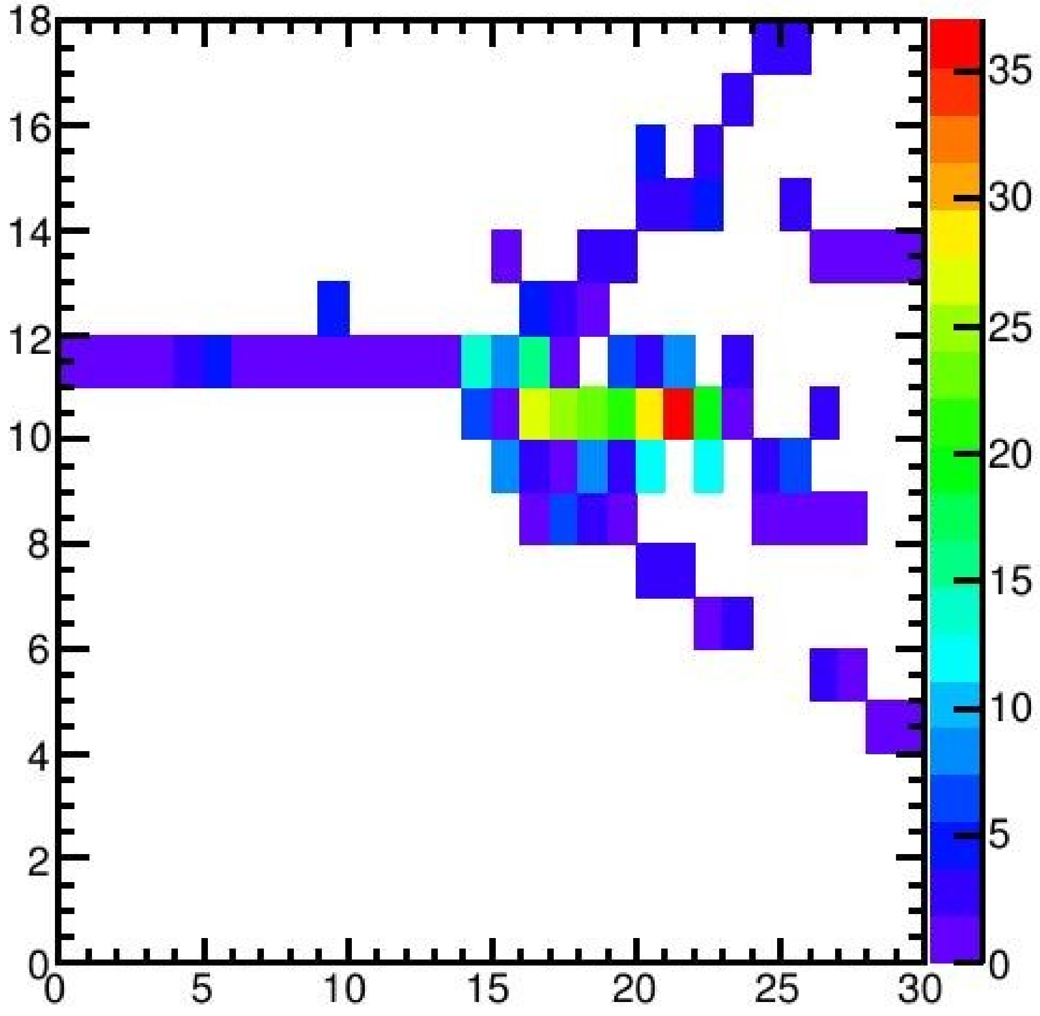
November 27, 2012

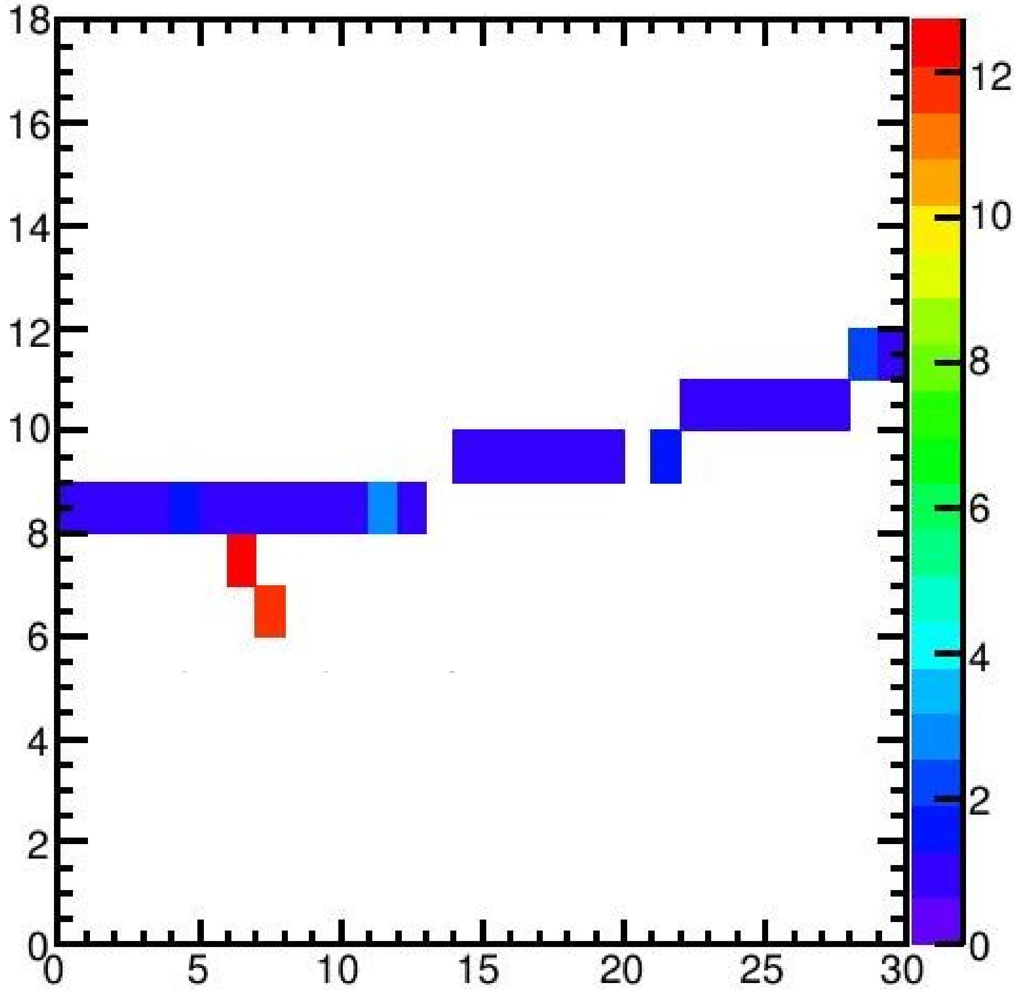
Outline

- Less about BigData, more about BigComputation
  - what professional inferrers and learners can do with large computational resources
- How to do inference once you have a model
  - likelihood-based inference by sampling
  - likelihood-free inference by simulation
- How to build models from scratch
  - deep learning: can the google cat be converted into a google boson?

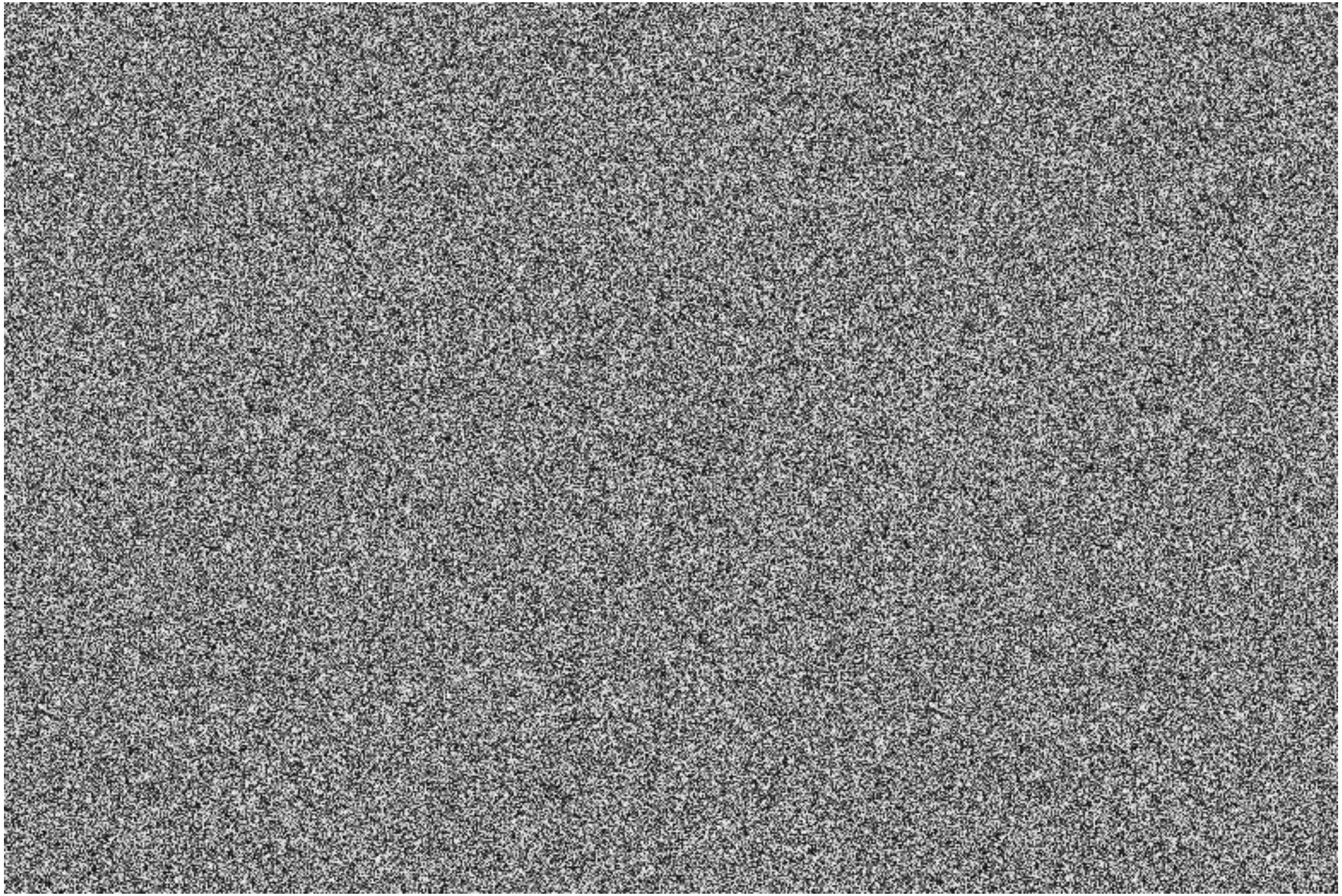
What is this?

What is this?

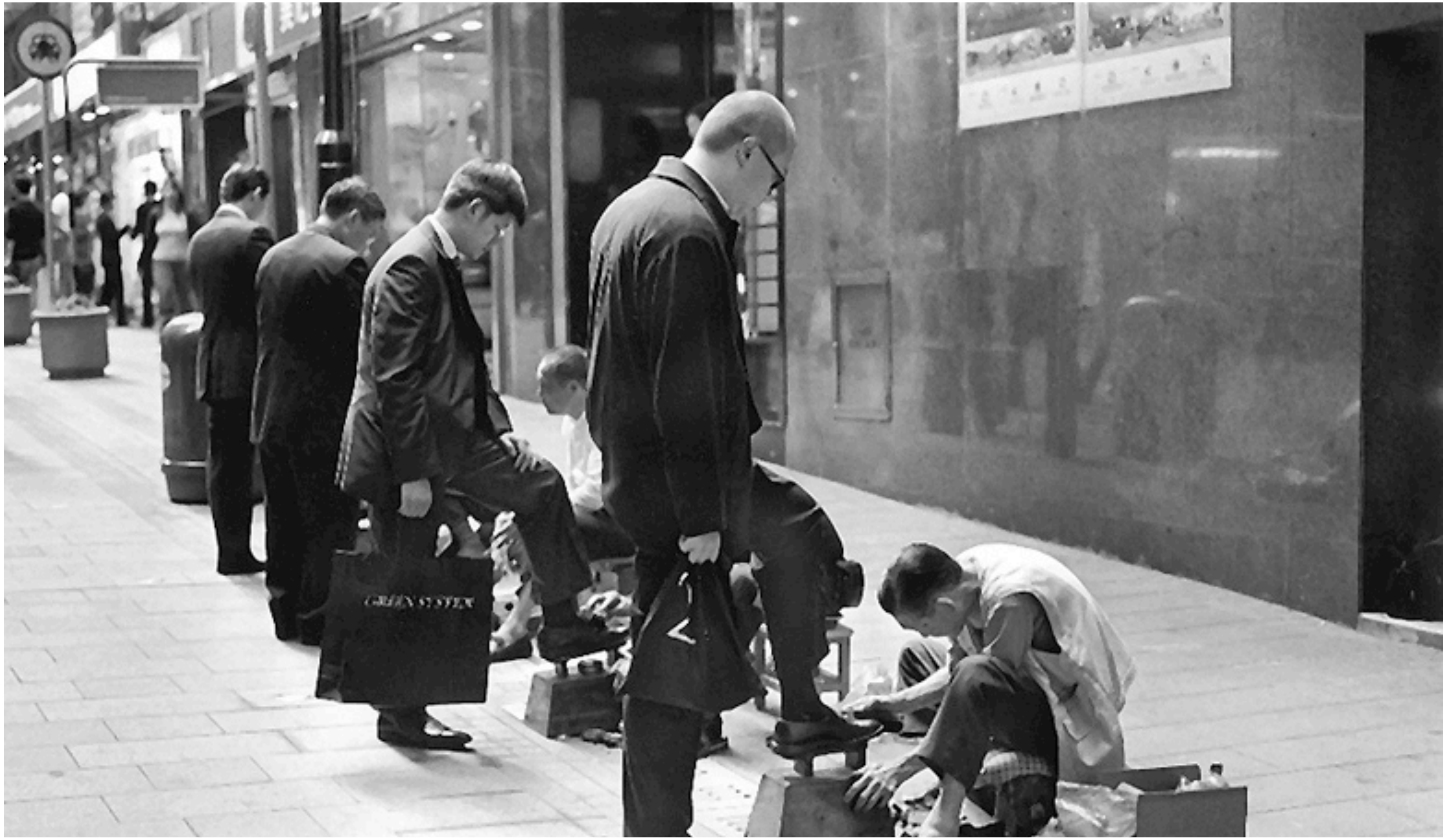

18
16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
30
35
30
25
20
15
10
5
0


18
16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
30
12
10
8
6
4
2
0

What is this?



What is this?

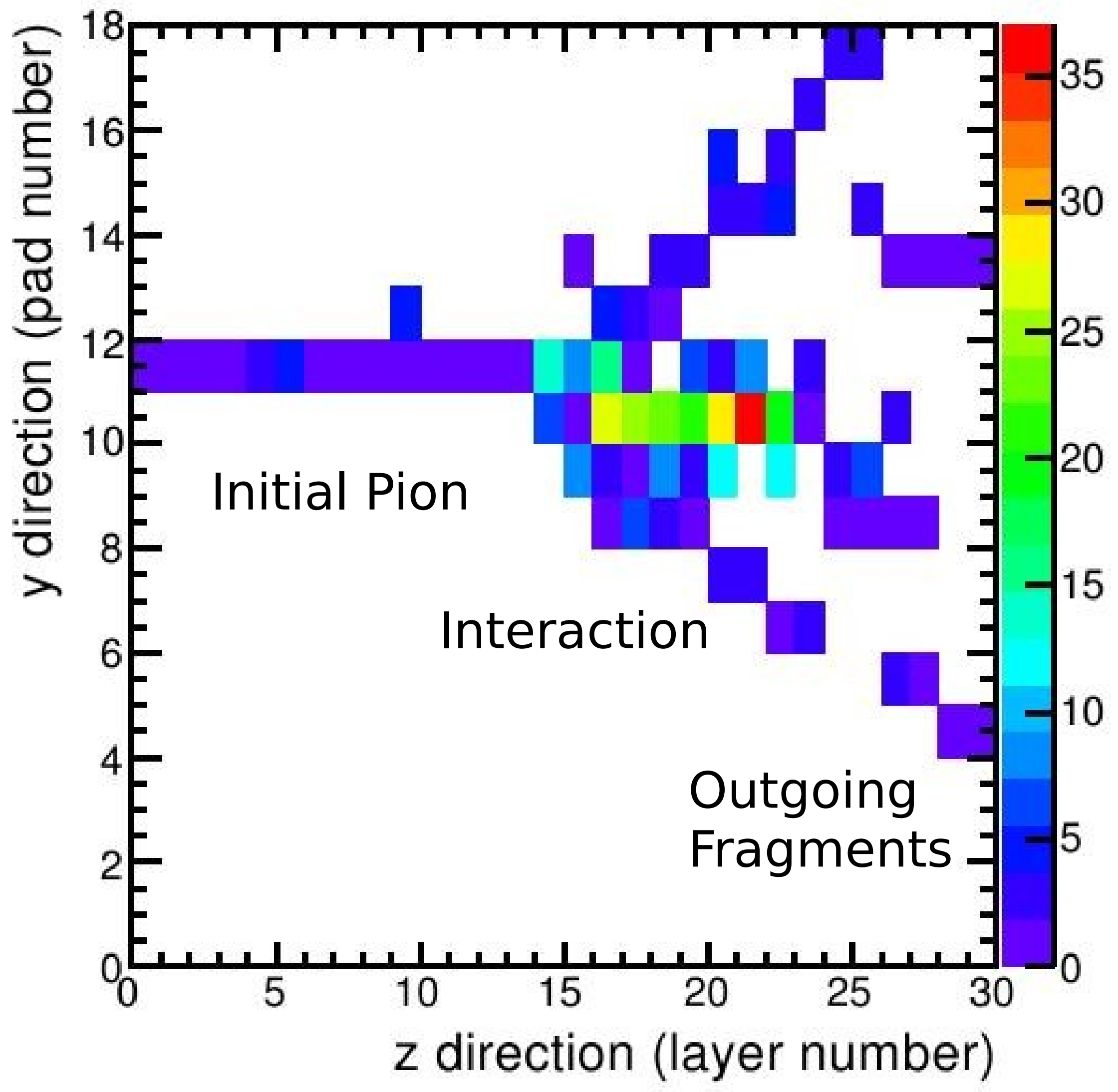


What is this?

Granularity and hadronic cascades

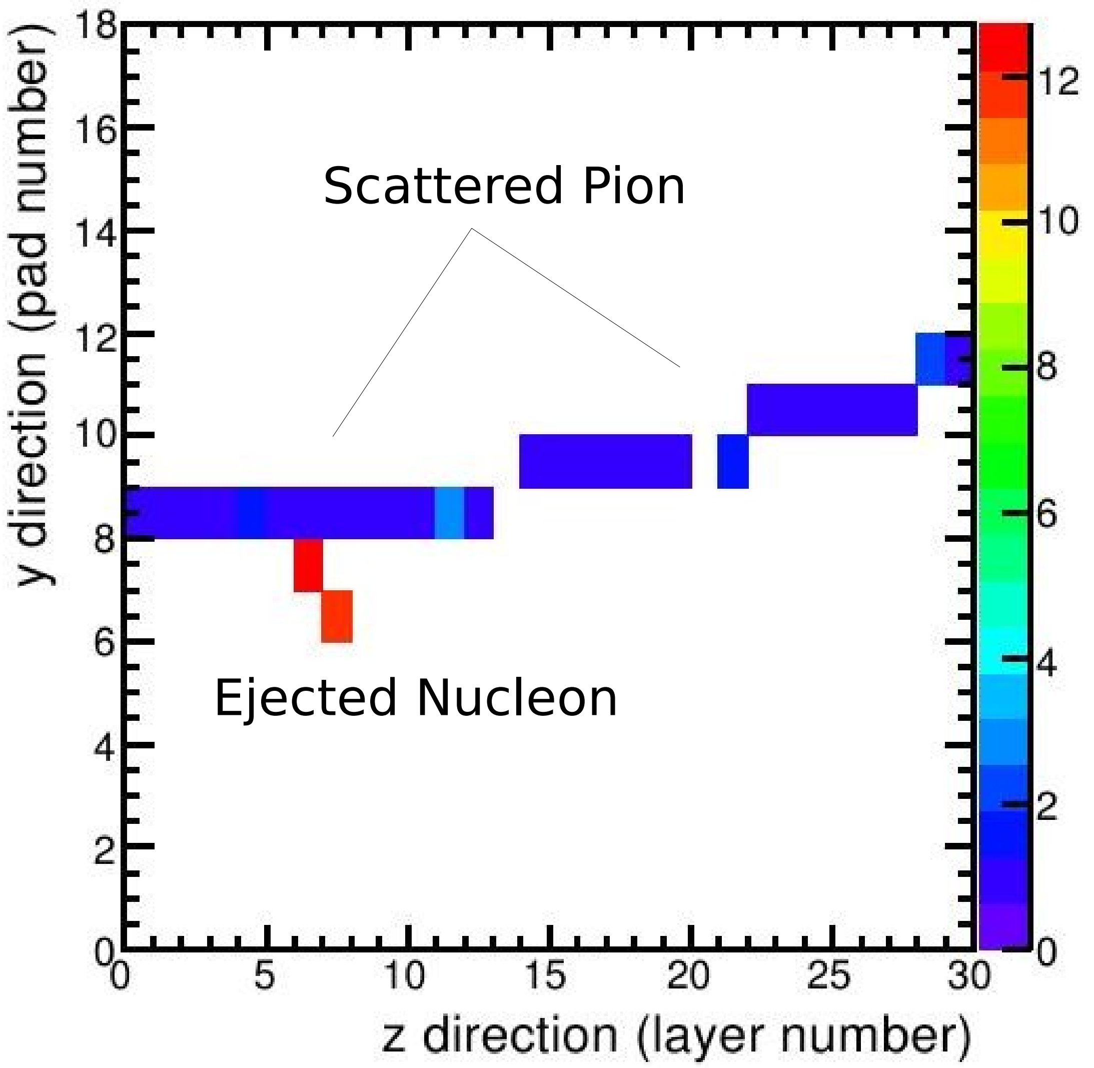
(Start of) Hadronic showers in the SiW Ecal

Complex and impressive



Inelastic reaction in SiW Ecal

Simple but nice



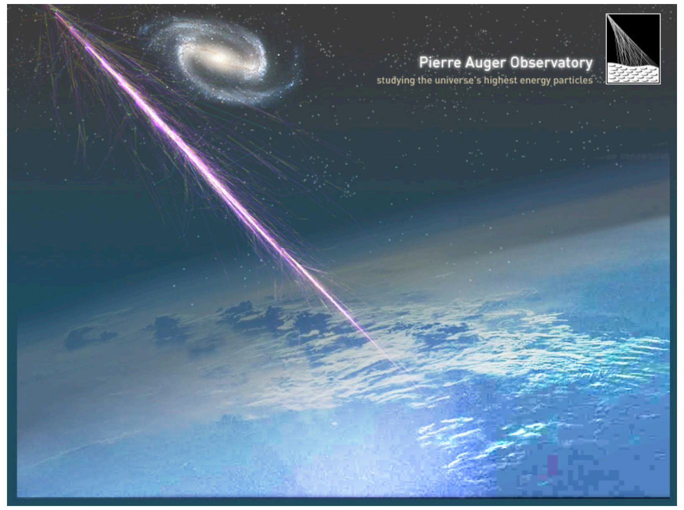
Short truncated showers

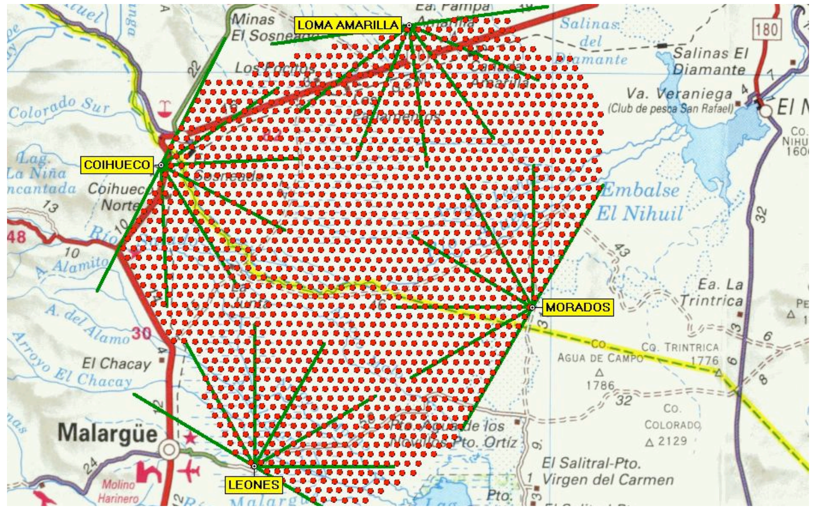
What is this?

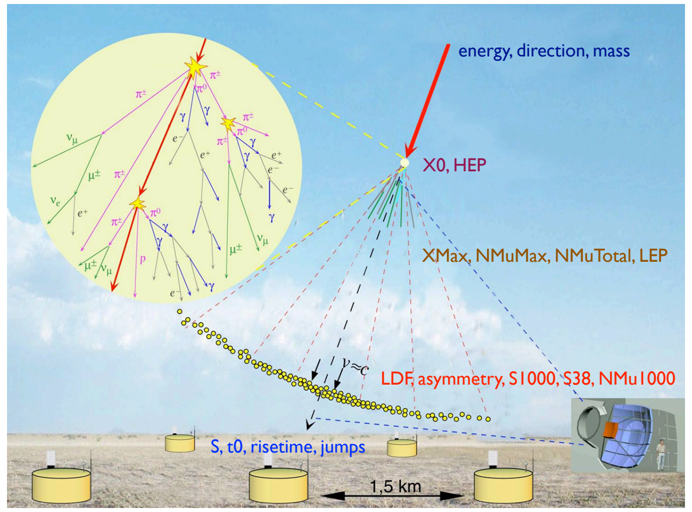
Models

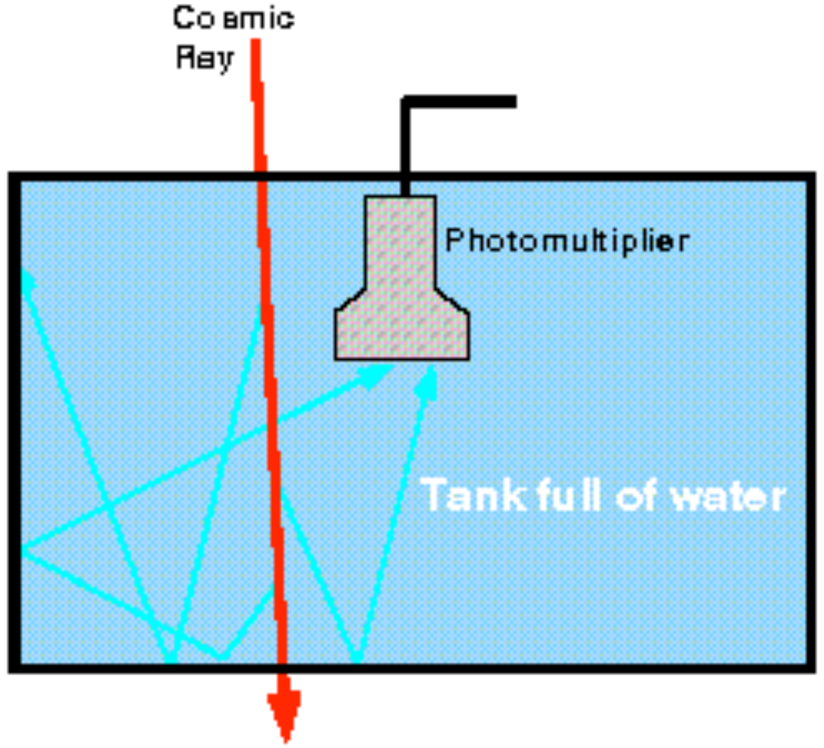
- Inference
  - if you want to be able to answer questions about observed phenomena, you need a model
  - if you want quantitative answers, you need a formal model
- Formal setup
  - x: observation vector, Θ: parameter vector to infer
  - likelihood: $p(x \mid \Theta)$
  - simulator: given Θ, generate a random x

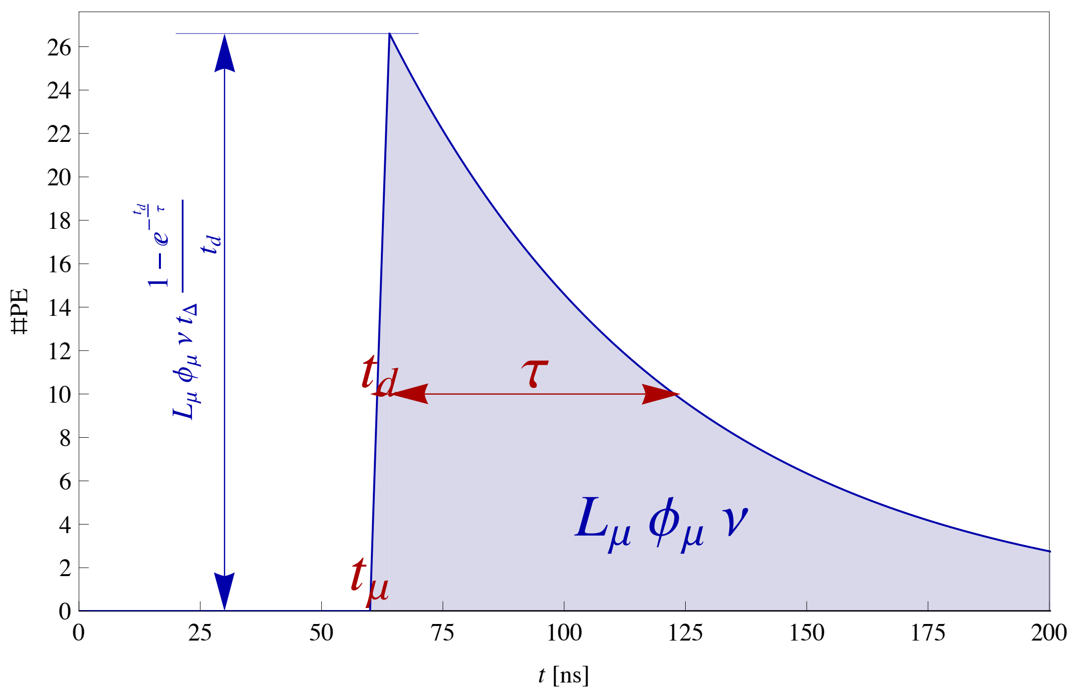
A formal model

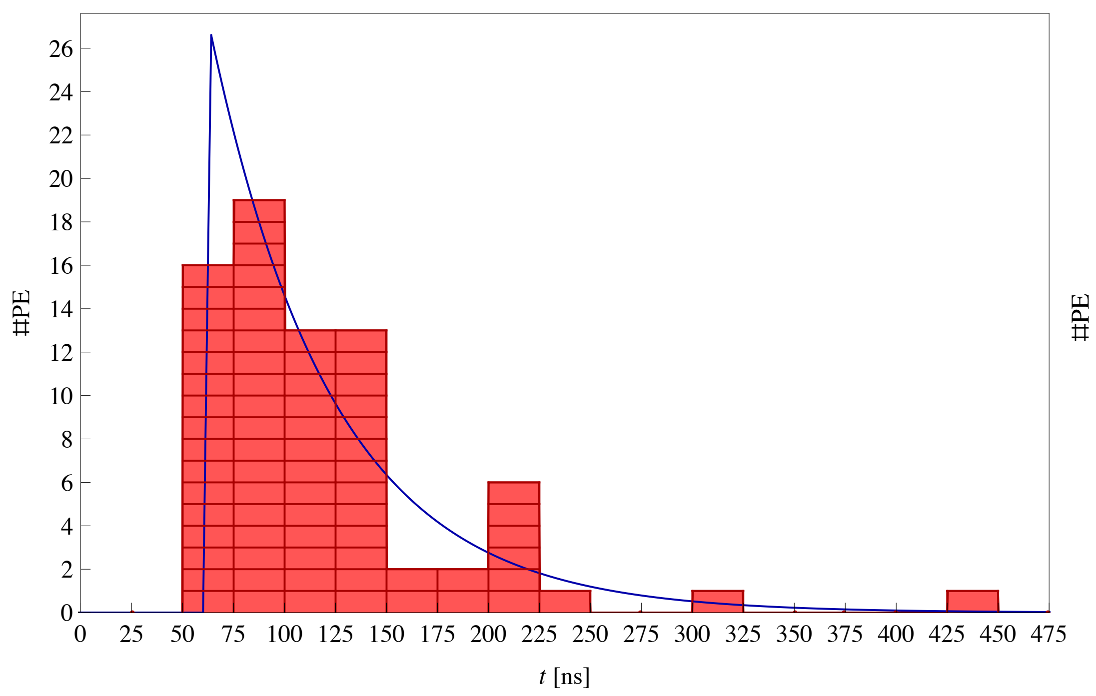


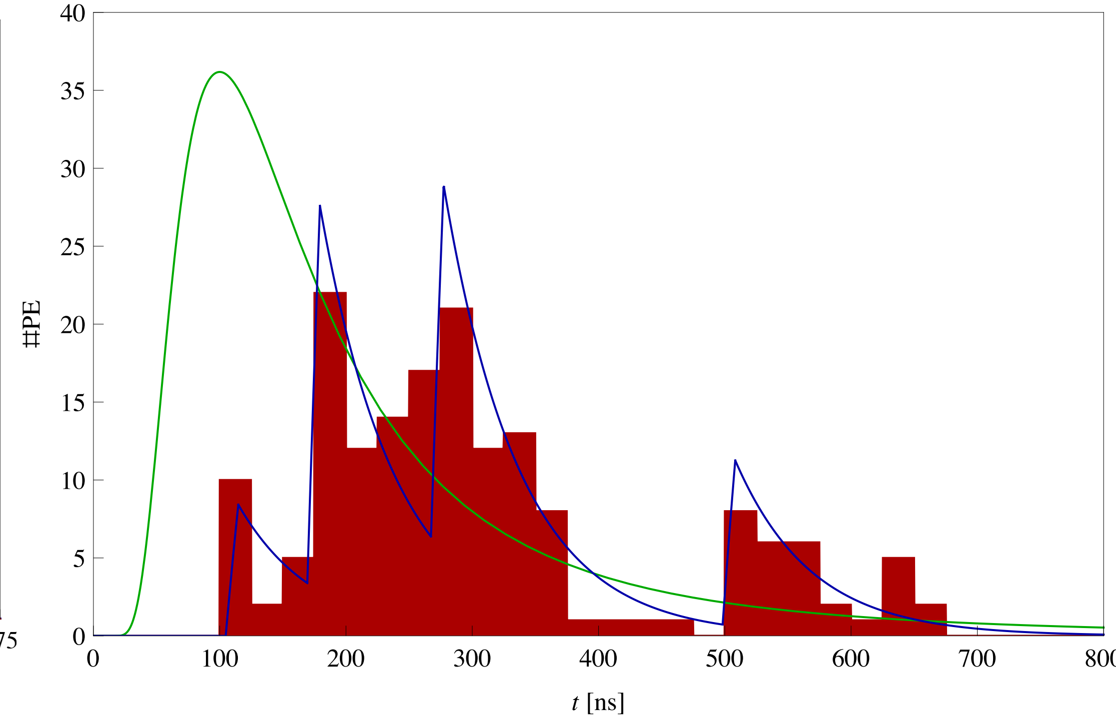












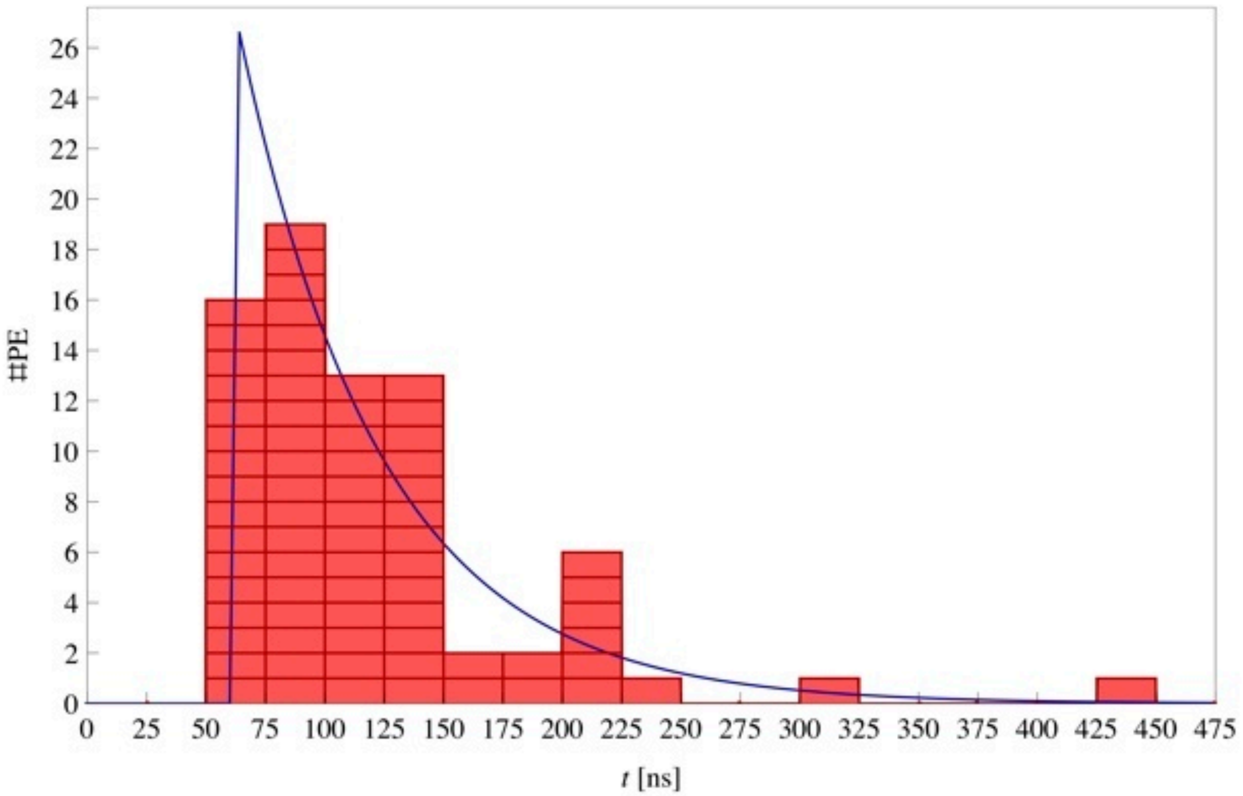
p(x | t)

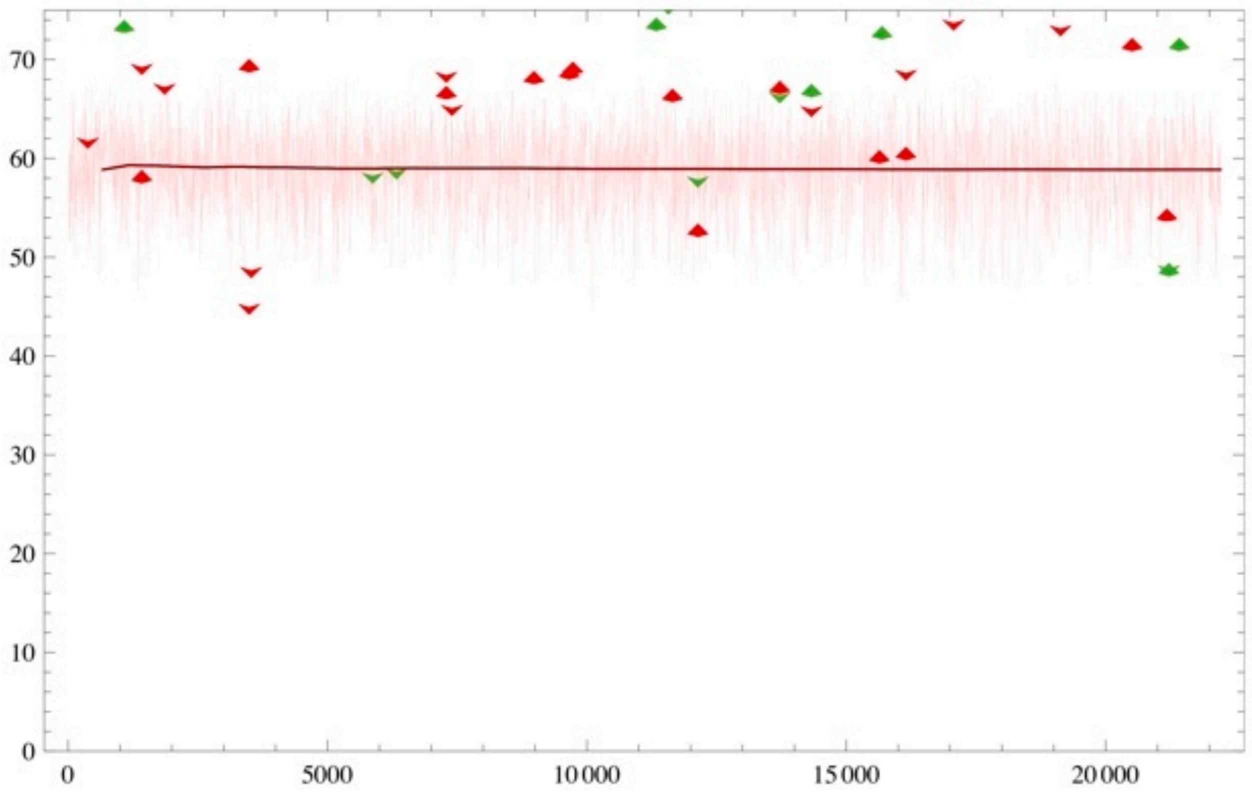
p(x | t1,...t4); p(t | Θ)

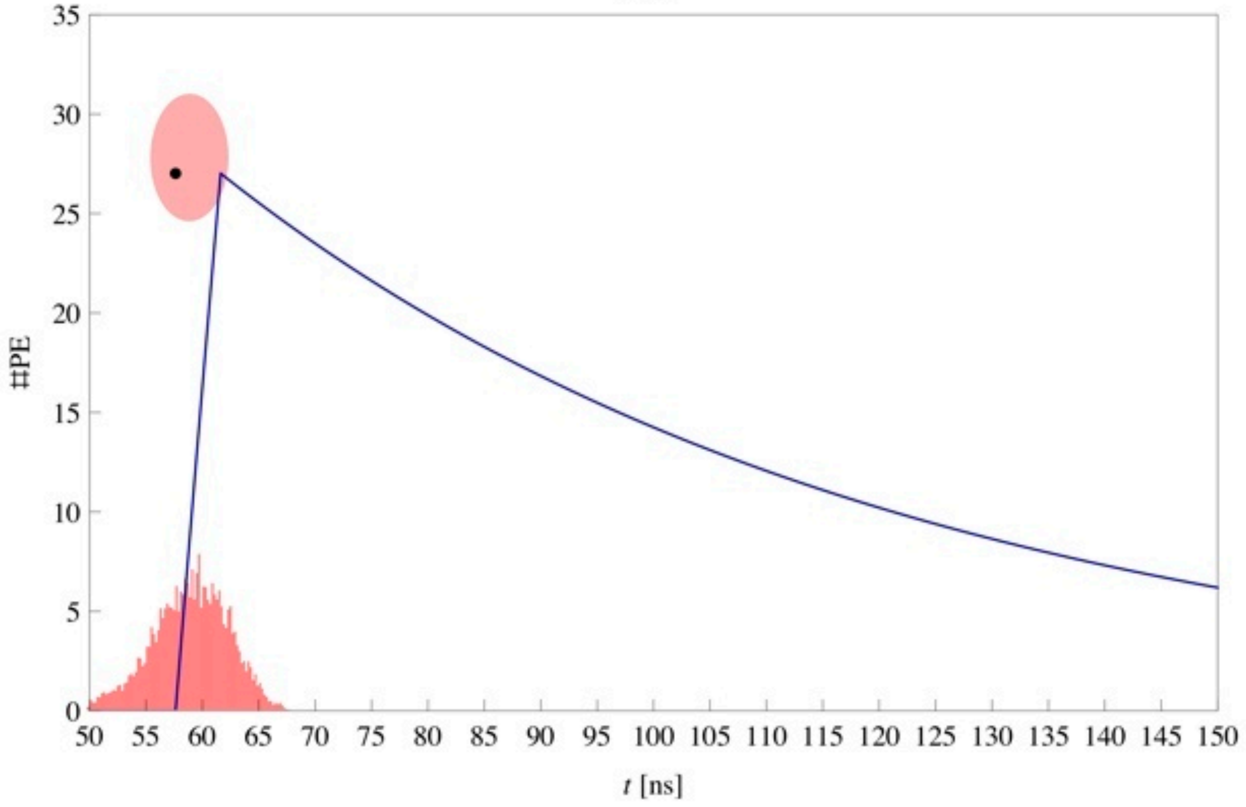
Inference by sampling

METROPOLIS$\left(p(x \mid \theta), p(\theta), \theta^{(0)}, \Sigma, T\right)$

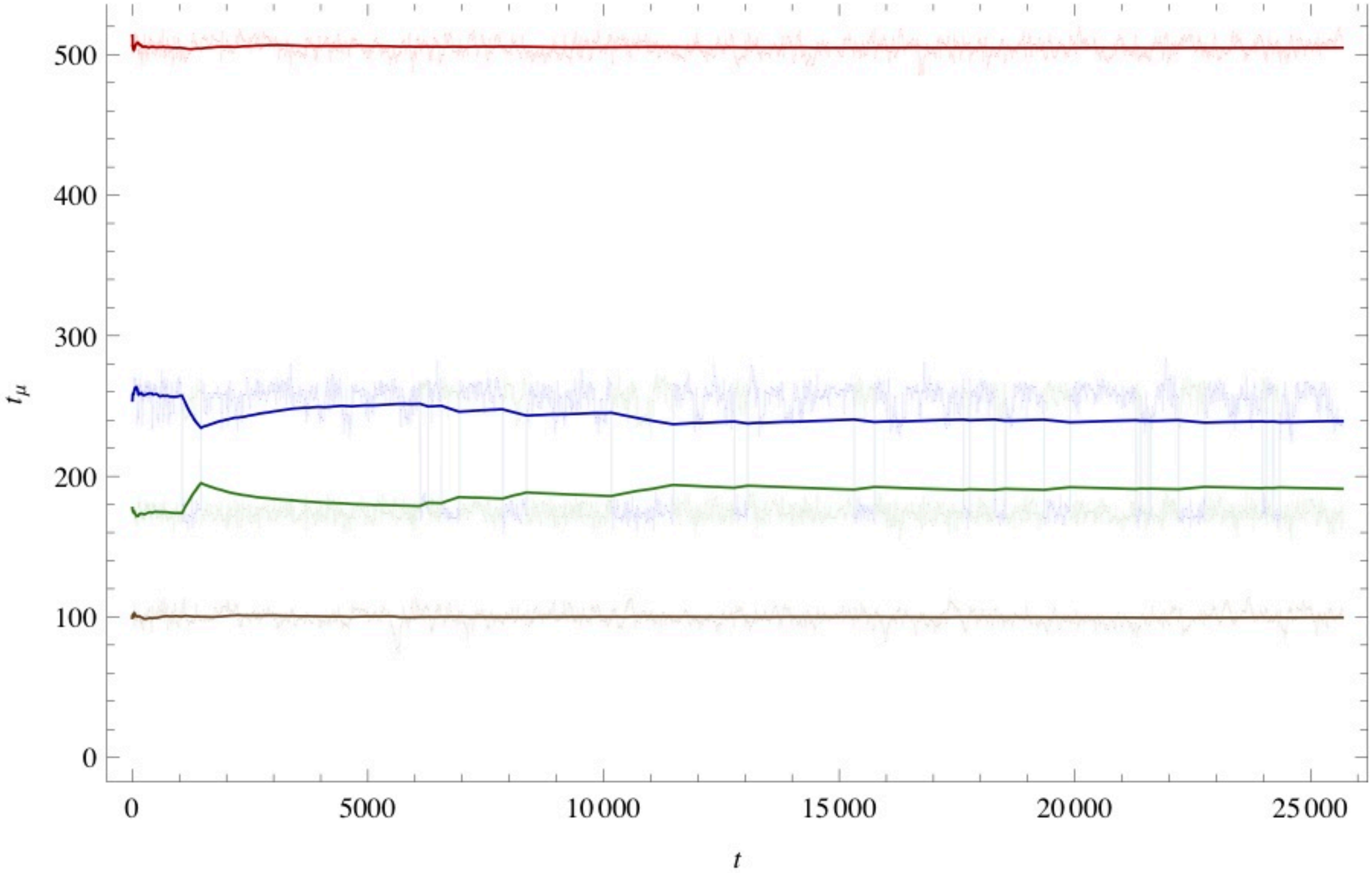
$\mathcal{S} \leftarrow \emptyset$
for $t \leftarrow 1$ **to** T
  $\theta \sim \mathcal{N}(\theta^{(t-1)}, \Sigma)$ ▷ *proposal*
  if $\frac{p(x \mid \theta)p(\theta)}{p(x \mid \theta^{(t-1)})p(\theta^{(t-1)})} > \mathcal{U}[0,1]$ **then**
    $\theta^{(t)} \leftarrow \theta$ ▷ *accept*
  else
    $\theta^{(t)} \leftarrow \theta^{(t-1)}$ ▷ *reject*
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{\theta^{(t)}\}$ ▷ *update posterior sample*
return $\mathcal{S}$







Inference by sampling



Inference by sampling

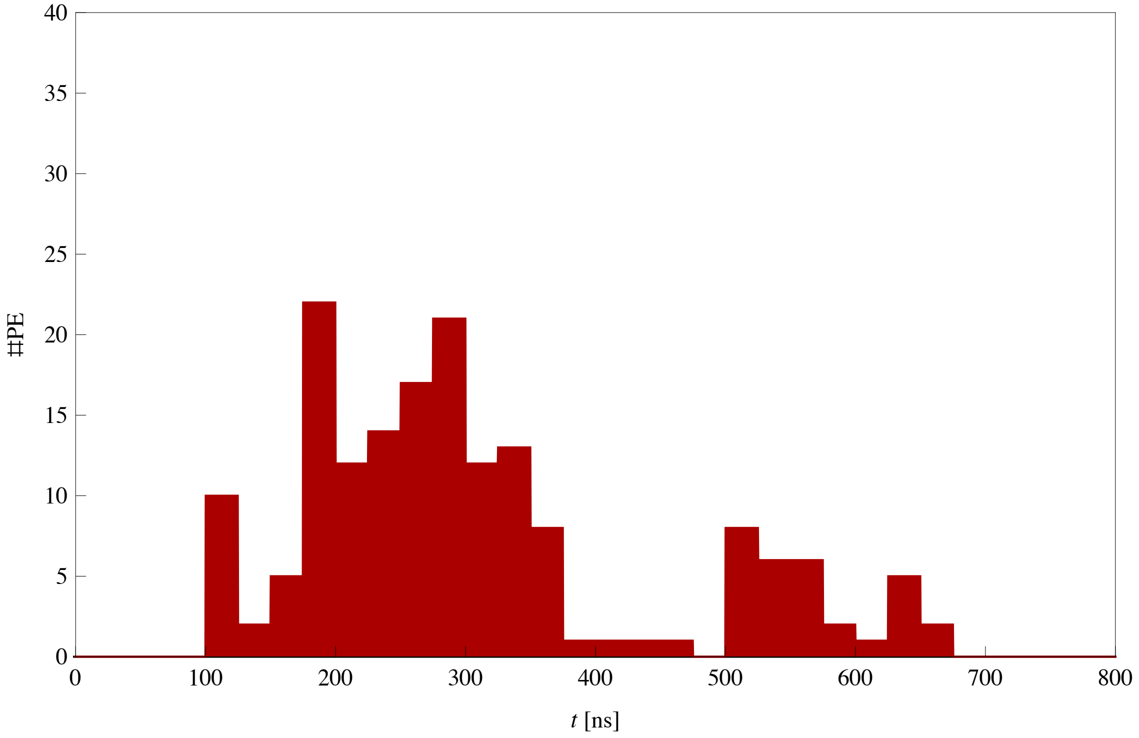
Inference by sampling

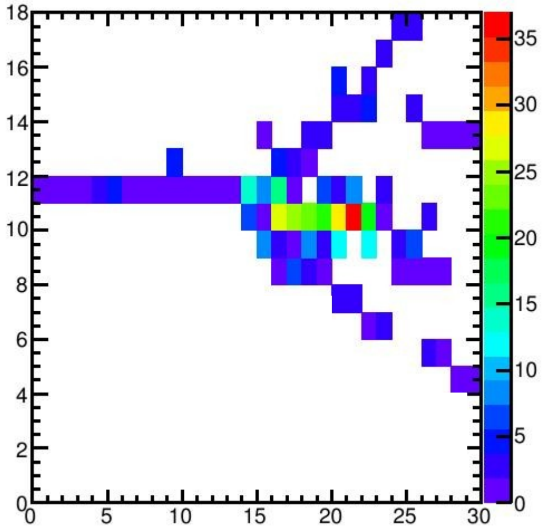
- An elegant approach for solving the inverse problem $p(x \mid \Theta) \rightarrow p(\Theta \mid x)$
  - needs the likelihood $p(x \mid \Theta)$ - a modeling requirement
  - needs to evaluate the likelihood a zillion times - a computational requirement
- When the likelihood cannot be computed but we can simulate from $p(x \mid \Theta)$
  - ABC: approximate Bayesian computation
  - only needs a similarity metrics between observables $K(x1,x2)$
  - needs to simulate from $p(x \mid \Theta)$ a zillion times - a computational requirement

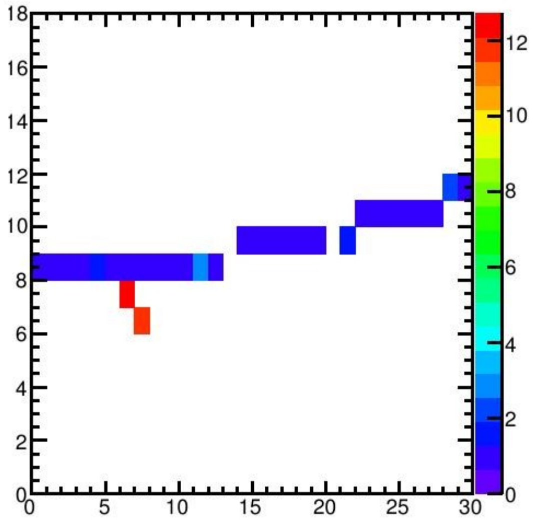
Outline

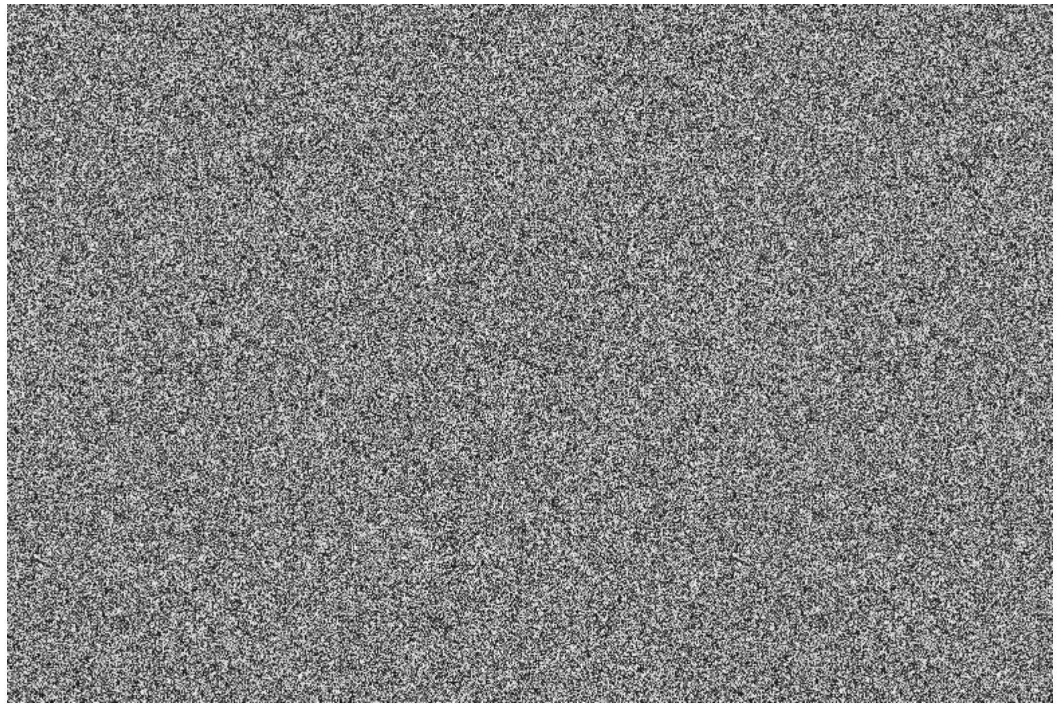
- Less about BigData, more about BigComputation
  - what professional inferrers and learners can do with large computational resources
  - of course related (BigData comes from BigComputers), but the viewpoint is different
- How to do inference once you have a model
  - simulation-based inference
- How to build models from scratch
  - deep learning: can the google cat be converted into a google boson?

How to build models for these?









The deep learning revolution

- Training multi-layer neural networks
  - biological inspiration: we know the brain is multi-layer
  - appealing from a modeling point of view: abstraction increases with depth
  - notoriously difficult to train until Hinton (stacked RBMs) and Bengio (stacked autoencoders), around 2006
  - the key principle is unsupervised pre-training
  - they remain computationally very expensive, but they learn high-level (abstract) features and they scale: with more data they learn more

The deep learning revolution

- The MNIST DBN

The deep learning revolution

- Google passes the "purring test"
  - 16K cores watching 10M youtube stills for 3 days
  - completely unsupervised

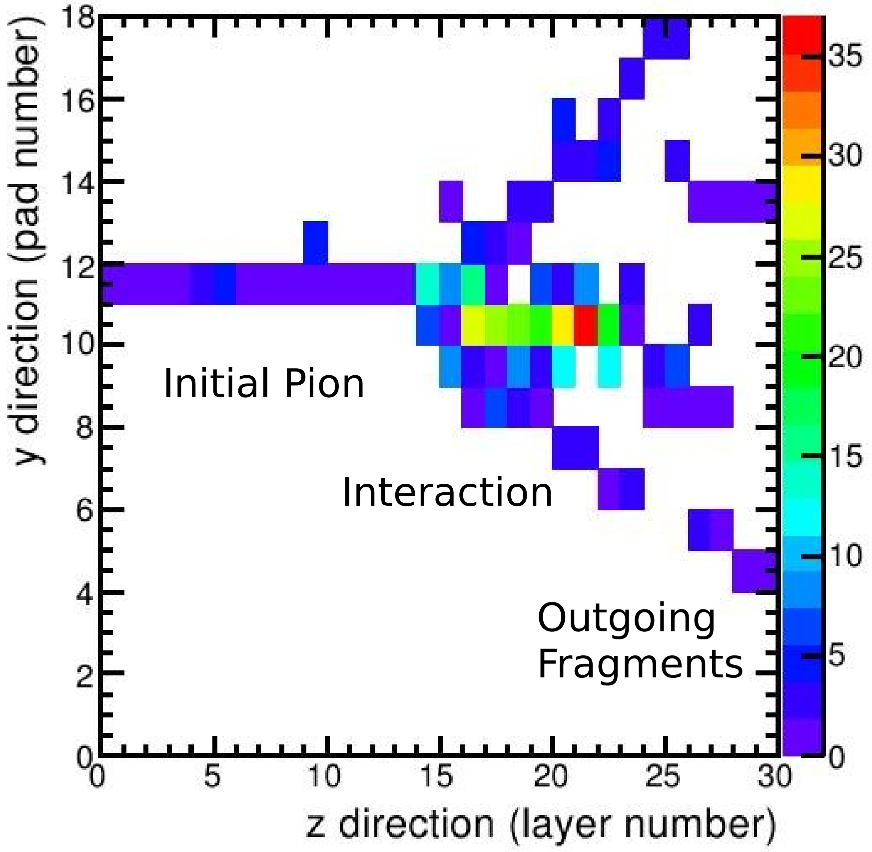
The deep learning revolution

- Can we also learn physics by observing natural (or man-designed) phenomena?

Granularity and hadronic cascades

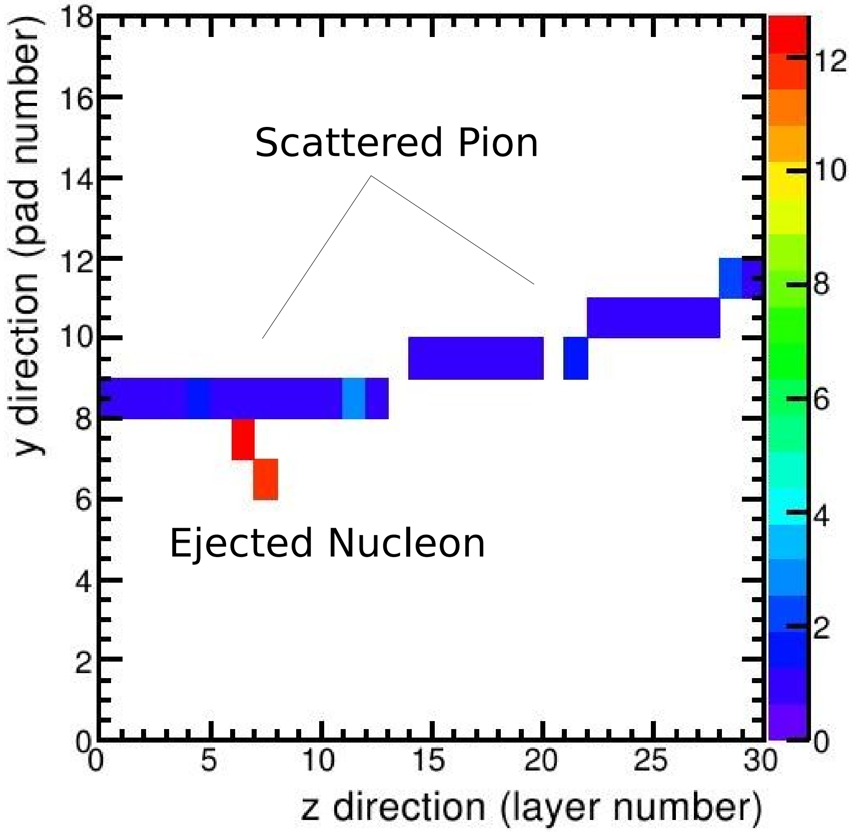
(Start of) Hadronic showers in the SiW Ecal

Complex and impressive



Inelastic reaction in SiW Ecal

Simple but nice



Short truncated showers